Using parametric B splines to fit specular reflectivities

N. F. Berk and C. F. Majkrzak
Materials Science and Engineering Laboratory, National Institute of Standards and Technology, Gaithersburg, Maryland 20899-0001

(Received 12 September 1994; revised manuscript received 5 January 1995)

Parametric B-spline curves offer a flexible and appropriate mathematical description of scattering length density profiles in specular reflectivity analysis. Profiles combining smooth and sharp features can be defined in low dimensional representations using control points in the density-depth plane which provide graded local influence on profile shape. These profiles exist in vector spaces defined by B-spline order and parameter knot set, which can be systematically densified during analysis. Such profiles can easily be rendered as adaptive histograms for reflectivity computation. B-spline order can be chosen to accommodate the aymptotic (large-Q) behavior indicated by reflectivity data. We describe an interactive fitting strategy in which the Nelder and Mead simplex method is used in the B-spline control point space to guide the discovery of profiles that can produce given reflectivity data. Examples using actual and simulated spectra are discussed.

I. INTRODUCTION

In this paper we describe the application of parametric B-spline curves to the model-free determination of neutron and x-ray scattering length density profiles from measurements of specular reflectivity, a problem of current and increasing interest.[1] The goal, briefly, is to determine a $\rho(z)$, representing the scattering length density of a thin film as a function of depth, from a spectrum $|R(Q)|^2$, representing reflectivity as a function of the wave vector or, equivalently, the incident or reflection angle. For a given profile $\rho(z)$, the specular reflectance $R(Q)$ can be computed exactly for practical purposes, but the inverse problem of determining $\rho(z)$ from $|R(Q)|^2$ is neither straightforward nor entirely well defined, the loss of phase information in relating $|R(Q)|^2$ to $R(Q)$ being the most evident—but not the only—mathematical obstacle. While the inverse scattering problem technically is solvable in one dimension when $R(Q)$ is known everywhere in Q,[2] the approach does not yet appear to be practical. Methods akin to isotopic substitution in crystallography have been shown to be effective in providing phase information in x-ray reflectivity from thin films[3] by varying the x-ray wavelength through the substrate absorption edge in order to selectively change atomic scattering lengths within the films. Maximum entropy methods—which are model-free, but in a different sense from the usage we employ below—also have been developed for obtaining density profiles from reflectivity spectra.[4,5]

Most approaches to analyzing reflectivities are variants of curve fitting[6–14] and employ the generic components illustrated in Fig. 1. These are the following: (1) Describe and (2) render (i.e., approximate) candidate density profiles; (3) compute corresponding trial reflectivities; (4) compare trials with reflectivity data, which entails the choice of an objective function, such as a χ^2; and (5) update profiles using an optimization or equivalent scheme to minimize the objective. In this context, the first stage, the mathematical description of profiles, is the one which mainly distinguishes model-based approaches from model-free methods. Model-based fittings test parametrized formulas, often resulting from or motivated by theoretical predictions.[6] Model-free descriptions test restricted function spaces. Since both approaches involve fit variables, distinguishing between them is not entirely clear-cut. The sharpest mathematical distinction is that models usually are nonlinear functions of their parameters; function spaces are spanned by linear com-

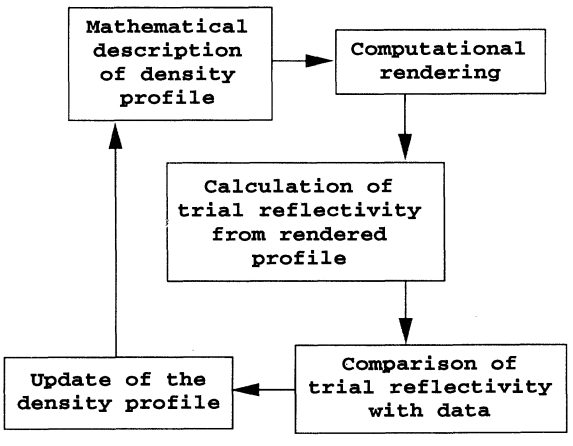

FIG. 1. Flow diagram for curve fitting approaches to specular reflectivity analysis.

binations of suitable bases. Consequently, models tend to probe limited classes of profile shapes, especially when using a small number of fit variables, while model-free descriptions, in principle, have a larger range.

In most model-free and model-based methods, candidate profiles of arbitrary shape are rendered into approximations suitable for computing the corresponding reflectivities.[15] Common rendering schemes are based on making rectangular approximations to $\rho(z)$, as in elementary numerical integration, since the reflectivity then can be computed exactly for the given rendering using transfer matrix or continued fraction techniques. In some model-free approaches, rectangular representations do double duty as the mathematical descriptions of candidate profiles, blurring the distinction between the first two components in Fig. 1. Such box function spaces usually must be high dimensioned in order to adequately represent smooth profiles, but effective searching of function spaces benefits from low-dimensional descriptions. The competing aspects of rectangular representations have been addressed in different ways in recent model-free analyses, including systematic dimensional refinement[9] and advanced Monte Carlo optimization techniques.[10,11] Here we will construct profiles as parametric B-spline curves, which admit reasonably low-dimensional descriptions of complex shapes, and render them as high-dimensional adaptive rectangular approximations for effectively exact computation of $|R(Q)|^2$ everywhere in Q.

Several recent works have employed nonrectangular representations of function spaces to describe scattering length density profiles for model-free reflectivity analysis.[7,8,12–14] Pedersen[7] and Pedersen and Hamley[12–14] exploited the utility of (nonparametric) cubic B-spline curves in this context, based on Glatter's original applications to small angle scattering.[16] Our use of parametric B-spline curves was arrived at independently[17] and was motivated in part by contact with literature on computer graphics, in which parametric B-spline curves and related curvilinear forms have received a great deal of attention. The concentration on parametric curves, or z-ρ trajectories, provides a fairly general approach to the description of profile shapes and their discovery by data fitting. Furthermore, the methods involve mathematical details we think are relevant and interesting, as will be discussed below.

B-spline theory may be found in several textbooks.[18–20] Nevertheless, in order to keep the present discussion reasonably self-contained and allow for a cohesive development, we give a focused review of parametric B-spline curves in the Appendix. In Sec. II we adapt this material to the analysis of reflectivity spectra and in Sec. III we describe our curve fitting methods in detail. Examples of applications are shown in Sec. IV. A summary and conclusion follow in Sec. V.

II. REPRESENTING DENSITY PROFILES

In this section we elaborate on the first box of Fig. 1, which we introduced in Sec. I. The remaining elements of that figure will be developed mainly in the next section. In model-free fitting methods, deciding on a mathematical description of $\rho(z)$ essentially means choosing a function space "large" enough to contain potentially interesting profiles but "small" enough to permit feasible surveys. Our response to these conflicting demands is to restrict candidates for $\rho(z)$ to parametric B-spline function spaces. This actually involves two separate mathematical notions. Parametric functions representing $\rho(z)$ are trajectories (general curves) in the z-ρ plane. Most z-ρ trajectories are not $\rho(z)$ functions, in fact, and thus are unacceptable as physical profiles. However, the same flexibility that permits "too much" variation also enables the description of $\rho(z)$ functions with features, such as corners and cusps, that are not easily provided otherwise. Thus the practical application of parametric curves to our fitting problem requires finding ways to sensibly restrict parametric flexibility without suppressing it all together. This means allowing the fitting procedure to probe near the boundary between physical and nonphysical profile shapes. The B-spline representation provides one way to handle this.

B splines are localized, piecewise continuous polynomial functions from which general splines can be constructed, and they serve as bases functions for suitably defined spline function spaces. For our purposes the effectiveness of B splines springs from a useful compromise between smoothness and locality. Constructing extended curves from localized bases usually entails losing some degree of smoothness (continuity of derivatives). The more localized the descriptions, the less smooth the curves, in general. At one extreme, rectangular representations of curves using box functions provide the sharpest spatial resolution but the least degree of smoothness; even continuity is lost. At the other extreme, Fourier representations give the smoothest possible descriptions but the least sharp spatial resolution or locality. Smoothness can be mimicked in rectangular representations only by using many boxes, while locality can be mimicked in Fourier representations only by using many frequencies. Generally, but not always, low-dimensional rectangular and Fourier representations are too restrictive for model-free applications. B splines, on the other hand, have mechanisms for adjusting the degree of smoothness, and thus locality, which can be adapted to various applications.

Parametric B splines are parametric curves in which the component coordinate functions are each represented as B-spline curves. The conjunction of these distinct elements provides a means of embedding physically meaningful $\rho(z)$ curves in moderately low-dimensional function spaces that prove to be at once rich and manageable.

A. Defining the space

Scattering length density profiles $\rho(z)$ may be represented by parametric B-spline curves of suitable order. We soon will choose cubic B splines, $k = 4$, but for a while we retain the generality of an arbitrary k. Using the conventions of the Appendix, Sec. 1, these curves

have the mathematical form

$$\mathbf{C}(u) = (z(u), \rho(u)) = \sum_{i=0}^{m} \mathbf{V}_i B_{i,k}(u) , \tag{1}$$

where u is the evolution parameter of the curve, the $B_{i,k}(u)$ are B-spline functions of order k, uniquely defined by a discrete set of parameter values $\{u_i\}_0^{m+k}$ called the knot sequence, and

$$\mathbf{V}_i = (z_i, \rho_i) , \quad i = 0, \ldots, m , \tag{2}$$

are the control points of the curve in the z-ρ plane, which for given order and knot sequence uniquely specify $\mathbf{C}(u)$.

As discussed in the Appendix, Sec. 2, the coordinate functions $z(u)$ and $\rho(u)$ belong to spline vector spaces $\mathbf{S}^k$ in which they vary independently. Thus the parametric curves $(z(u), \rho(u))$ belong to direct product spaces $\mathbf{S}^k \otimes \mathbf{S}^k$.[21] Unrestricted direct product spaces are too large for the problem at hand, however. In addition to all the curves which may describe physically meaningful $\rho(z)$, they contain many that cannot. These are reentrant z-ρ trajectories (having loops, etc.), which are not single-valued functions.[22] The geometrically relevant elements of $\mathbf{S}^k \otimes \mathbf{S}^k$ are trajectories for which the position coordinates $z(u)$ are nondecreasing functions of u to ensure that z values do not recur along the curves. Let us call this portion of the space $\mathcal{S}^k$. Then we have

$$\mathbf{S}^k \otimes \mathbf{S}^k \supset \mathcal{S}^k \supseteq \mathbf{S}^k . \tag{3}$$

The second relation in Eq. (3) follows from the fact that (for $k > 0$) $\mathbf{S}^k$ is equivalent to a subspace of $\mathbf{S}^k \otimes \mathbf{S}^k$ in which $z(u)$ is an affine tranformation of u, which also makes it a subspace of $\mathcal{S}^k$. To produce monotonic $z(u)$, it suffices to restrict $\{z_i\}$ to monotonic sequences:

$$z_0 \le z_1 \le \cdots \le z_m . \tag{4}$$

The proof is simple. Nondecreasing $z(u)$ requires a nonnegative first derivative $z^{(1)}(u)$. The derivative of

$$z(u) = \sum_{i=0}^{m+1} z_i B_{i,k}(u) \tag{5}$$

is

$$z^{(1)}(u) = (k-1) \sum_{i=\delta-k+2}^{\delta} \frac{z_i - z_{i-1}}{u_{i+k-1} - u_i} B_{i,k-1}(u) , \tag{6}$$

for $u_\delta \le u < u_{\delta+1}$.[23] The denominators in Eq. (6) are never negative, and B splines are everywhere nonnegative, Sec. 1 of the Appendix; so the ordering in Eq. (4) is sufficient to make $z^{(1)}(u) \ge 0$, as required. For $k > 2$, strict ordering is not necessary for monotonic behavior, however. If $z_i < z_{i-1}$ for some i, the sum in Eq. (6) may still be non-negative. Higher-order spline curves tend to be stiffer in this regard; a larger k entails more terms in the sum, providing more opportunities for compensation of a negative contribution. Thus the boundary of $\mathcal{S}^k$ is convoluted and difficult to characterize, but functions on this boundary are important descriptively. They provide instances of $\rho(z)$ in which overall smoothness is punctuated by sharp features, such as edges and corners. It is desirable that model-free fitting methods be able to find them. The constraint of Eq. (4) is overly restrictive but relatively easy to impose. We describe a strategy for doing this in Sec. III A.

It is desirable to be able to refine the description of a curve by adding more control vertices without initially perturbing it. Remarkably, this is possible;[20,24] a B-spline curve can be exactly rerepresented by a longer control sequence. Refinement follows naturally from the vector space properties of B splines;[18–20,25] any spline curve may be embedded in spaces of higher dimension. A general solution of this problem is provided by the "Oslo" method.[24,25] The basic idea is to choose one or more curve segments for refinement and subdivide the corresponding parameter knot segments. The Oslo algorithm then computes a new B-spline basis on the larger knot sequence and finds the expanded control graph needed to reconstruct the original curve in the new basis. The Oslo algorithm is defined in the Appendix, Sec. 2.

B. Choosing B-spline order

The degree of smoothness required of functions intended to represent physical scattering length density profiles is difficult to pin down and will depend, among other concerns, on the range and quality of the reflectivity data being analyzed. Typically, we expect model density profiles to be more or less smooth, with occasional interruptions by sharp features, possibly at the edges of the film and at interior interfaces.[26] One of the few formal quides to smoothness requirements in this context is provided by the Riemann-Lebesque-Lightill theorem,[27] which relates the asymtotic (large-Q) behavior of a Fourier transform to the integrability of the direct function and its derivatives. It is well known, (see, e.g., Refs. 28, 29) that as $Q \to \infty$, the specular reflectivity $|R(Q)|^2$ is given exactly by the Born approximation

$$|R(Q)|^2 \sim \frac{16\pi^2}{Q^2} |\rho[Q]|^2 , \tag{7}$$

where

$$\rho[Q] = \int_{-\infty}^{\infty} e^{iQz} \rho(z)\, dz \tag{8}$$

is the Fourier transform of $\rho(z)$, and Q is the scattering wave vector. According to the Riemann-Lebesgue-

Lighthill theorem, if $\rho^{(n)}(z)$ is square integrable, then $\rho[Q]$ falls to zero at least as fast as Q^{-n-1}, for $Q \to \infty$. The case $n = 0$, meaning square-integrable $\rho(z)$, conforms to the original Riemann-Lebesque lemma. If $\rho^{(s)}(z)$ is the highest derivative that is square integrable, then $\rho[Q]$ falls to zero no faster than Q^{-s-1}. The limit $s \to \infty$ corresponds to "smooth" functions in the formal sense; then $\rho[Q]$ falls to zero faster than any power law (e.g., the Fourier transform of a Gaussian). If r denotes the highest derivative which is continuous, then $s = r + 1$. Thus, if C^r is the (highest) continuity class of $\rho(z)$, then $\rho[Q]$ falls no faster than Q^{-r-2}. Therefore, for $\rho(z) \in \mathbf{S}^k \otimes \mathbf{S}^k$, $|R(Q)|^2$ falls no faster than Q^{-2k-2}. The case $k = 1$ (Q^{-4}) corresponds nominally to Fresnel reflection from sharp interfaces. However, the description of sharp interfaces does not require the use of $k = 1$ B splines. As discussed in the Appendix, k collinear control vertices produce a straight segment, and these can be vertically aligned to form an edge. Similarly, $k = 2$ (linear) spline curves are not required for tentlike corners. These can be formed by $k-1$ coincident control vertices. In general, parametric B-spline representations of $\rho(z)$ can display isolated features that are sharper than the nominal smoothness dictated by B-spline order.

The limiting degree of smoothness required of $\rho(z)$ is set effectively by the fastest anticipated fall of the reflectivity, in concert with other, harder-to-quantify considerations. For example, if the data exhibit a Fresnel-like (Q^{-4}) fall at large Q, $\rho(z)$ must contain at least one sharp edge or step of signigicant height; it is likely, however, that a smoother overall description than that afforded by $k = 1$ is needed for a low-dimensional model. The observed fall at large Q is allowed by a description based on B splines of higher order. Similarly, an asymptotic Q^{-6} requires at least $k = 2$, but a low-dimensional linear model may not be correct, overall. On the other hand, there would seem to be little gain in using an unnecesarily high-order description. Aside from the heavier computational burden this entails, it also means that more control vertices have to be involved in the description of sharp features, and it forces tighter, longer-ranged correlations among spline segments.

In this work, we use cubic B splines, $k = 4$, for the parametric descriptions of $\rho(z)$, which would appear to offer an effective compromise of these competing demands. In particular, $k = 4$ accommodates an asymptotic fall as strong as Q^{-10}, which is fast compared to typical reflectivity data, but just fast enough for density profiles having everywhere continuous curvature. Each segment of the curve needs four control vertices, and no more than four consecutive segments interact. In this description, true verticals and flats need four collinear vertices, and true cusps require threefold vertex multiplicitites. The effects of rendering, discussed below, act to weaken these exacting requirements.

C. Fixing end points

We assume that the end points of $\rho(z)$ are known and fix them by constraining the vertices at each end of the control graph. These dependent vertices are known as "phantoms" to the convey the idea they may also be introduced behind the scenes.[20,30] Let the candidate films extend from a front face at $z = 0$ to a back face at $z = L$, where L is the film thickness. If in fact L is not known with reasonable accuracy, or cannot be inferred from multilayer diffraction peaks, different values must be tried. We do not use L as a dynamic fit variable. For $k = 4$ and an arbitrary knot sequence, we have derived the phantoms for the curves of Eq. (1) as

$$\hat{\mathbf{V}}_0 = \frac{(0, \rho(0)) - a_1 \mathbf{V}_1 - a_2 \mathbf{V}_2}{a_0} \tag{9}$$

and

$$\hat{\mathbf{V}}_m = \frac{(L, \rho(L)) - b_1 \mathbf{V}_{m-1} - b_2 \mathbf{V}_{m-2}}{b_0}\,, \tag{10}$$

where the coefficients are

$$\begin{aligned} a_0 &= \frac{(u_4 - u_3)^2}{(u_4 - u_1)(u_4 - u_2)}\,, \\ a_1 &= \frac{(u_3 - u_1)(u_4 - u_3)}{(u_4 - u_1)(u_4 - u_2)} + \frac{(u_5 - u_3)(u_3 - u_2)}{(u_5 - u_2)(u_4 - u_2)}\,, \\ a_2 &= \frac{(u_3 - u_2)^2}{(u_5 - u_2)(u_4 - u_2)}\,, \end{aligned} \tag{11}$$

and

$$\begin{aligned} b_0 &= \frac{(u_{m+1} - u_m)^2}{(u_{m+3} - u_m)(u_{m+2} - u_m)}\,, \\ b_1 &= \frac{(u_{m+3} - u_{m+1})(u_{m+3} - u_m)}{(u_{m+3} - u_m)(u_{m+2} - u_m)} \\ &\quad + \frac{(u_{m+1} - u_{m-1})(u_{m+2} - u_{m-1})}{(u_{m+2} - u_{m-1})(u_{m+2} - u_m)}\,, \\ b_2 &= \frac{(u_{m+2} - u_{m+1})^2}{(u_{m+2} - u_{m-1})(u_{m+2} - u_m)}\,. \end{aligned} \tag{12}$$

We assume, further, that the end-point values of $\rho(z)$ are given by the scattering length densities of the incident and substrate media. Apparant deviations from this behavior are easily accounted for by control vertices near the end points. These constraints, therefore, do not prove to be overly restrictive and can be changed for particular applications.

On a given knot sequence, a candidate $\rho(z)$ thus is defined by a control graph

$$\hat{\mathbf{V}}_0, \mathbf{V}_1, \ldots, \mathbf{V}_{m-1}, \hat{\mathbf{V}}_m\,, \tag{13}$$

which comprises a subgraph visible to the fitting procedure,

$$\mathcal{V} = \mathbf{V}_1, \mathbf{V}_2, \ldots, \mathbf{V}_{m-2}, \mathbf{V}_{m-1}\,, \tag{14}$$

and the phantom vertices $\hat{\mathbf{V}}_0$ and $\hat{\mathbf{V}}_m$, which are computed with Eqs. (9) and (10). Considerations germane to B-spline knot selection are discussed in the Appendix, Sec. 4. We initialize $\rho(z)$ on uniform parameter knot sequences, Eq. (A15), with $k = 4$. If we let N be the

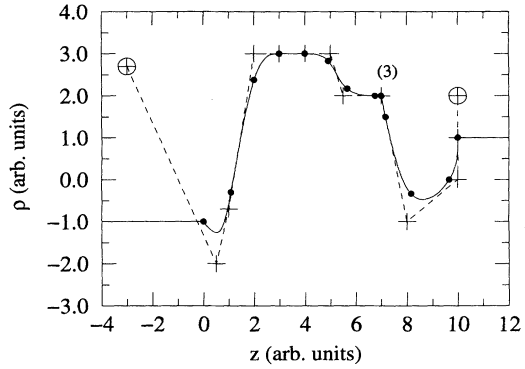

FIG. 2. Fictitious scattering length density profiles made with parametric uniform cubic B splines. The curves consists of 12 visible segments divided by 13 knots (•). The first and last of these are fixed points at (ρ, z) coordinates (0,−1) and (10,1). The flat portions to the left and right are not part of the spline. The curve is defined by 13 visible control vertices (+) and 2 phantoms (⊕) calculated to fix the ends. The line connecting the vertices is the control graph. Four collinear control vertices, 3–6, produce a flat portion in segment 4. Triply degenerate vertices, 8–10, produce a corner cusp at $z = 7$. The curve ends with a vertical slope at (10,1).

number of visible control vertices and M the number of knots on the curve, then for this case $N = M$. Figure 2 illustrates some basic anatomy of parametric B-spline scattering length density profiles which are relevant for applications. Our B-spline computational methods are outlined in the Appendix, Sec. 3.

III. FITTING REFLECTIVITY DATA

A. Searching the space

We use the Nelder-Mead simplex method[31–33] to search the z-ρ plane for visible control graphs $\mathcal{V}$ which lead to acceptable agreement between calculated and measured reflectivities. A simplex in n-dimensional Cartesian space is a polyhedron having $n + 1$ vertices. Since we have preempted the term "vertex" for use with parametric spline curves, we refer to simplex vertices, henceforth, as nodes. A simplex is formed by joining every node to every other node. When the n edges incident to a node are linearly independent, the simplex spans the n-dimensional space. In the simplex method the nodes are vectors holding the fit variables. The dimension of the space depends on how one chooses to group these variables. For parametric B-spline curves it is natural to consider individual control vertices, $\mathbf{V}_i = (z_i, \rho_i)$, as being two-dimensional fit variables. A node $\mathcal{V}$ then constitutes an $(m-1)$-dimensional vector in the simplex space, with each of its components (i.e., slots in the list of fit variables) holding a control vertex. A simplex comprises m such nodes, $\mathcal{V}_q$, for $q = 0, \ldots, m-1$. Each node is a visible control graph, which, when augmented by phantoms, generates a candidate $\rho(z)$ for the fit. Each $\rho(z)$, in turn, generates a value of an objective function, a function (actually, a functional) which measures the closeness of the fit. The nodes of a simplex can be ordered according to their objective values, from lowest to highest.

The Nelder-Mead algorithm is a set of finite linear transformations on simplices,[32] which are designed to seek lower values of the objective by systematically driving the simplex away from high values. Each move starts with an attempt to replace the highest node by a lower node, not necessarily a new lowest. The discovery of a new lowest node triggers an attempt to stretch the simplex in the successful direction; failure to improve on the highest node triggers a move which compresses the simplex toward the current lowest node. In all, some 12 different moves are available in the standard method. The algorithm does not terminate automatically. If the lowest node is a local minimum (and the entire simplex is within its basin), the simplex contracts toward it until stopped. However, the lowest node is never replaced by a higher one, a property which endows the simplex with a degree of stability and persistence. The moves are scaled by the size of the simplex, and so the simplex tends to move more "slowly" when it is near a local minimum or in states in which it is difficult to improve on the high node. In worst cases, the simplex contracts exponentially without reaching a minimum, which may be the method's most serious deficiency. This problem can be mitigated by expanding the simplex from time to time about the low node.

Since the simplex moves are linear transformations, they preserve linear relationships among nodal components which are common to all the nodes. As Nelder and Mead originally pointed out,[31] this manifold property can be used to advantage to impose constraints on the simplex search. For example, a repeating component pattern common to all nodes is a linear invariant[34] and thus is preserved by simplex moves. As we will demonstrate in Sec. IV, this provides a handy way of incorporating multilayer structures into parametric B-spline descriptions of $\rho(z)$.

The manifold property cannot strictly enforce the ordering of the $\{z_i\}$ indicated in Eq. (4), because an ordering constraint is not a linear invariant.[35] Thus a simplex that comes near the hard border represented by Eq. (4) may cross it and ultimately produce unphysical candidates for $\rho(z)$. Strict ordering can be imposed with an explicit sorting operation, but this sometimes will be incompatible with end-point control; sorting only the visible sequence $\{z_i\}_1^{m-1}$ does not guarantee that the dependent phantoms will be consistent with the sort, while sorting the entire sequence $\{z_i\}$ may change the end points. Alternatively, simplex moves producing out-of-sort nodes could be rejected, but this would require expensive testing and decisions about what moves to substitute without defeating the carefully designed simplex strategy. For simplicity, in the applications to be described in Sec. IV, we sorted the visible $\{z_i\}$. Our experience, however, is that this precaution mainly comes into play in early stages of fitting, before the simplex has been well trained on the objective. Figure 3 illustrates

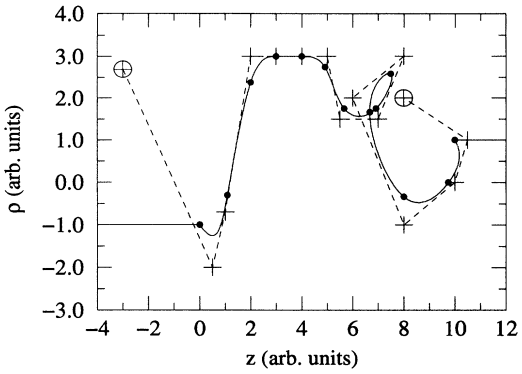

FIG. 3. The unphysical loop results from lifting the degeneracy in Fig. 2 with out-of-sort vertices. The unphysical bow at the right edge results from vertex 13 being too far to the right. In practice such loops are rarely observed because of the initialization method used, but bowed edges are common if the specified film thickness is too small.

behavior in the unphysical portion of the spline space.

For the objectives to be minimized we use ratio-sensitive functionals[11] of the type

$$F[\rho(z)] = \frac{1}{N}\sum_{n=1}^{N}\left|\frac{D_n - E_n}{D_n \wedge E_n}\right|^p , \tag{15}$$

where D_n is the measurement of $|R(Q)|^2$ at $Q = Q_n$, E_n is the corresponding trial value for a candidate $\rho(z)$, N is the number of data points, and $D_n \wedge E_n$ is the smaller of D_n and E_n. The exponent p is an integer. D and E are everywhere non-negative; in fact, each lies in the interval $[0,1]$. This objective class comprises symmetric, non-error-weighted measures of closeness of fit, punishing bad ratios of D and E equally, regardless of whether $D > E$ or $D < E$. Ratio measures are well suited to comparing reflectivity curves, which typically range over many orders of magnitude.[36] We have found that $p = 1$ works well with the simplex search method, since it is effective in reducing the rate of simplex shrinkage. For this case, the objective can be expressed as

$$F[\rho(z)] = \frac{1}{N}\sum_{n=1}^{N}\frac{D_n \vee E_n}{D_n \wedge E_n} - 1 , \tag{16}$$

where $D_n \vee E_n$ is the larger of D_n and E_n. The first term in Eq. (16) is just the average upper-to-lower ratio of the two curves.

B. Rendering and computation

For specular scattering from a film having infinite extents in the x and y directions and a scattering length density which varies only along the z direction, it is necessary to solve the one-dimensional Schrödinger equation

$$\psi''(z) + k_z^2(z)\psi(z) = 0 , \tag{17}$$

where

$$k_z^2(z) = k_{0z}^2\left[1 - \frac{4\pi\rho(z)}{k_{0z}^2}\right] = k_{0z}^2 n(z)^2 . \tag{18}$$

In Eq. (18), $k_{0z} = k\sin\theta$, where θ is the angle of incidence of the neutron beam on the film, measured from the film surface, and where $\hbar k$ is the momentum of the incident beam. In conventional experiments, the film being studied is sandwiched between fronting (incidence) and backing media (one of which typically is air) having constant $\rho(z)$, with $z = 0$ locating the film's front edge and $z = L$ the back. Then, the wave function in the fronting medium is

$$\psi(z) = e^{ik_f z} + Re^{-ik_f z} , \tag{19}$$

while in the backing medium,

$$\psi(z) = Te^{ik_b z} , \tag{20}$$

where $k_f = k_z(z)$ for $z < 0$, and $k_b = k_z(z)$ for $z > L$. In these equations, $R = R(Q)$ and $T = T(Q)$ are the reflection and transmission amplitudes, respectively, and $Q = 2k_{0z}$. Analytical solutions of this problem are not available, generally, and so a degree of approximation is needed to compute $R(Q)$ for arbitrary $\rho(z)$. We employ a common solution method in which $\rho(z)$ is rendered in a rectangular approximation. First, the interval $[0, L]$ containing the film is subdivided by a mesh of discrete z values, $\{Z_m\}$, and then $\rho(z)$ is replaced by

$$\rho(z) \to \rho(Z_m) , \quad Z_m \le z < Z_{m+1} , \tag{21}$$

which constitutes a set of slabs of constant scattering length density. In our case, we cut the slabs $[Z_m, Z_{m+1}]$ by uniformly subdividing each parameter knot interval $[u_i, u_{i+1}]$ during B-spline rendering. The slabs thus vary in thickness along the z axis according to the local shape of $\rho(z)$. We solve the Schrödinger equation exactly for the slab-rendered $\rho(z)$ by using the transfer matrix method.[37,38] This leads to the equation

$$\begin{pmatrix} T \\ in_b T \end{pmatrix} = \mathsf{M}\begin{pmatrix} 1+R \\ in_f(1-R) \end{pmatrix} , \tag{22}$$

where $n_f = n(z)$ for $z < 0$, and $n_b = n(z)$ for $z > L$,

$$\mathsf{M} = \mathsf{M}_M\mathsf{M}_{M-1}\cdots\mathsf{M}_2\mathsf{M}_1 , \tag{23}$$

for M slabs, and

$$\mathsf{M}_m = \begin{pmatrix} \cos\delta_m & \frac{1}{n_m}\sin\delta \\ -n_m\sin\delta_m & \cos\delta_m \end{pmatrix} , \tag{24}$$

with $\delta_m = k_{0z}n_m(Z_{m+1} - Z_m)$, and $n_m = n(z)$ for $Z_m \le z < Z_{m+1}$. Equation (22) is then solved for $R(Q)$.

For finite Q these solutions become exact in the limit that $M \to \infty$, but for finite M, rectangular rendering is intrusive for the reasons discussed in Sec. II B. Scattering by a stack of slabs ultimately produces Fresnel behavior for $Qd > 1$, where d is the largest slab thick-

ness, regardless of the shape of the candidate $\rho(z)$. Thus, a rendered $\rho(z)$ which fits data up to $Q = Q_{\max}$ for a choice of M generally will not fit for $Q > Q_{\max}$ using the same M. However, for a given $Q_{\max}$, there is a smallest $M = M_{\min}$, which must be empirically determined, such that using $M > M_{\min}$ will not visibly improve the quality of fit. In general, of course, it is desirable to fit data using large M, to mitigate the effects of rectangular rendering, while also using the smallest number of fit variables that can work in order to facilitate the search for acceptible $\rho(z)$. In the current method, the descriptive and rendering requirements on $\rho(z)$ are uncoupled, as in model-based methods. We can search a relatively low-dimensional but rich function space without having to exact rendering penalties over a given, possibly large, range of data.[39]

IV. EXAMPLES

We illustrate the methods that have been described here with two applications to actual neutron reflectivity data and one to simulated neutron data. Thus from this point on we consider only the determination of real-valued $\rho(z)$. We have found it effective to start fits with as few control vertices as can allow reasonable agreement with low-Q data. Profiles thus trained often can be coerced into good fits with the aid of Oslo refinement and the device of "trickling" data into the fit, one or a few points at a time. The converse appears not to be true; profiles initially trained by high-Q data generally do not lead to good fits. In our experience each set of reflectivity data presents its own challenges to analysis, which seem best met by trial-and-error experimentation and a readiness to quickly abandon unproductive strategies. The fitted spline profiles we show below typically are rendered with 20–30 slabs per parameter knot segment. In all cases we check that the rendering is unobtrusive over the range of data used. The fitted profile curves shown in the figures are plotted with straight-line interpolation of the rendered spline points.

Figure 4 shows a fit to neutron reflectivity data from a titanium oxide film.[40] The film thickness is 160 Å. In the experiment the fronting medium is silicon and the backing medium is an electrolytic aqueous solution.[41] This fit employs seven visible control vertices on a uniform parameter knot sequence. The figure also shows the fit[40] obtained from a least-squares model refinement method[42] in which the scattering length density was described as a collection of constant-density slabs joined by error function segments. The reflectivities from these two fits lie close to each other except in the noisy portion of the spectrum at large Q. This difference may reflect the fact that the model fit was error weighted, thus emphasizing the higher reflectivity values (smaller error bars) in the noisy portion at the expense of the lower values. The ratio objective function used in the spline fit, as mentioned in Sec. III A, is not error weighted. The model fit also incorporates instrumental resolution while the spline fit does not. Note that both are unable to fit the data accurately at very low Q, but they miss by the same amount. Clearly there is a systematic effect in these data that neither method captures. The two fitted density profiles have similar shapes.

Figure 5 presents two parametric B-spline fits to the same data in Fig. 4, each using six control vertices and uniform knot sequences. The calculated reflectivities are nearly indistinguishable from each other over the range of data. The shape of the profile drawn with the solid line is roughly similar to the spline profile in Fig. 4, but the profile shown with the dotted line is substantially different. These two profiles, including the fronting and backing, are approximately related by an inversion in the z-ρ plane about a point (z_c, ρ_c) near the center of the film, where $z_c = L/2$ and $\rho_c = [\rho(0) + \rho(L)]/2$. This operation is not an exact symmetry of specular reflectivity but it is known to be a symmetry of the Born approximation.[5] Specifically, if $\bar{\rho}(z)$ is the inversion of $\rho(z)$, then

$$\bar{\rho}(z) = 2\rho_c - \rho(2z_c - z) \tag{25}$$

everywhere in z, and its Fourier transform, for real valued $\rho(z)$, is

$$\bar{\rho}[Q] = 2\rho_c\delta(Q) - e^{2iQz_c}\rho[Q]^*; \tag{26}$$

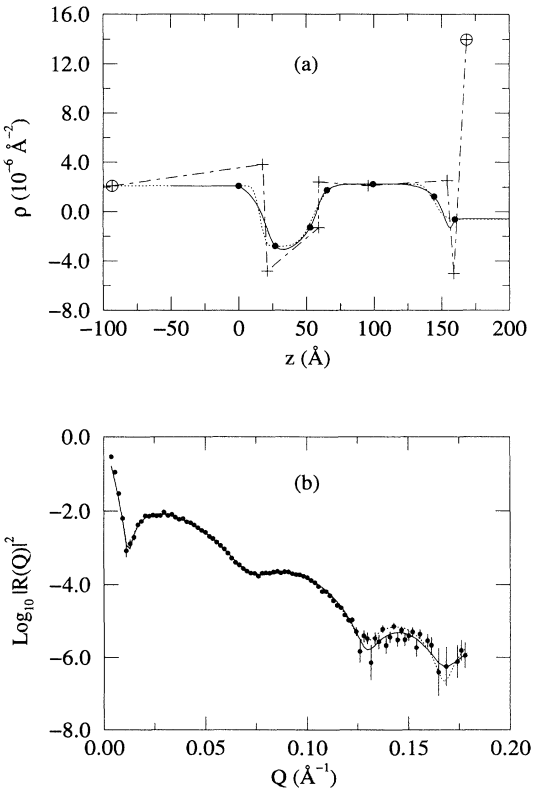

FIG. 4. Parametric B-spline fit of reflectivity from a thin titanium oxide film in contact with electrolyte. (a) Spline profile using six segments, with symbols as in Fig. 2. The dotted line is the model profile from Ref. 40. (b) Reflectivity data (points), spline fit (solid line), and model fit from Ref. 40 (dotted line).

so for $Q \neq 0$,

$$|\bar{\rho}[Q]|^2 = |\rho[Q]|^2 , \tag{27}$$

independently of the center of inversion, (z_c, ρ_c). Thus from Eq. (7), the two profiles produce the same reflectivities, asymptotically. Since the spectrum for the titanium oxide film at small Q effectively lies beyond the critical region where the Born approximation necessarily breaks down, finding an approximately inverted transform of $\rho(z)$ may not be surprising. It is interesting, nonetheless, that inverted forms were discovered rather easily in the course of exploring different spline fitting strategies for the spectrum. In this case, however, the inverted shapes can be rejected on physical grounds.[40]

Figure 6 displays nine B-spline fits to the data of Fig. 4 employing fitting strategies that resulted in different numbers of control vertices and various uniform and nonuniform parameter knot segments, ranging from 5 to 15 visible segments. Inverted shapes, in the sense of Fig. 5, were found for several of these. The corresponding reflectivities lie close to each other over this Q range, notwithstanding the detailed differences in the density profiles. However, these profiles are generally similar to each other and physically meaningful. For example, the relatively sharp feature seen in some of the curves near $z = 0$ could be interpreted as suggesting a thin layer of silicon oxide, which often is present on silicon substrates. Evidently these data do not require the clear presence of a silicon oxide feature in the profile, but the group of fits on the whole shows that such a feature is consistent with the spectrum. To this limited but useful extent, the reflectivity measurement determines a class of acceptable scattering length density profiles for the titanium oxide film.

Figure 7 shows a parametric B-spline fit to neutron reflectivity data from a multilayer consisting of seven layers of deuterated polystyrene-protonated polyisoprene diblock copolymer on silicon, with air as the fronting medium.[43] The thickness of the multilayer is approximately 1420 Å. The fit was accomplished with 42 control vertices, corresponding to 6 vertices per layer, and a uniform parameter knot sequence. A key ingredient in achieving this fit with relative ease was to train the Nelder-Mead simplex to be aware of the approximately periodic multilayer structure without having to touch the fitting algorithm. We did this by initializing the fit with random control graphs, each consisting of seven nonoverlapping groups of six vertices. The manifold property of the Nelder-Mead algorithm, Sec. III A, effectively hon-

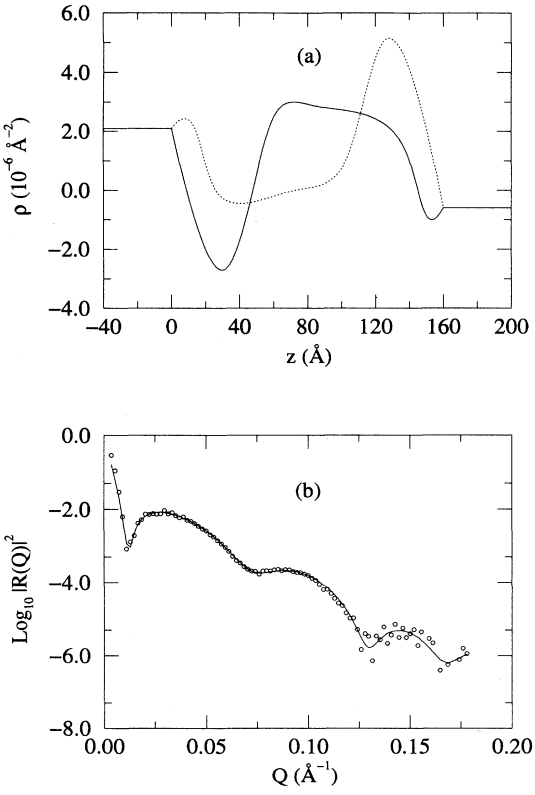

FIG. 5. Additional fits to data in Fig. 3. (a) Two spline profiles using five segments. (b) Reflectivites shown by corresponding solid and dotted lines, which effectively coincide, and the data (circles) as in Fig. 3.

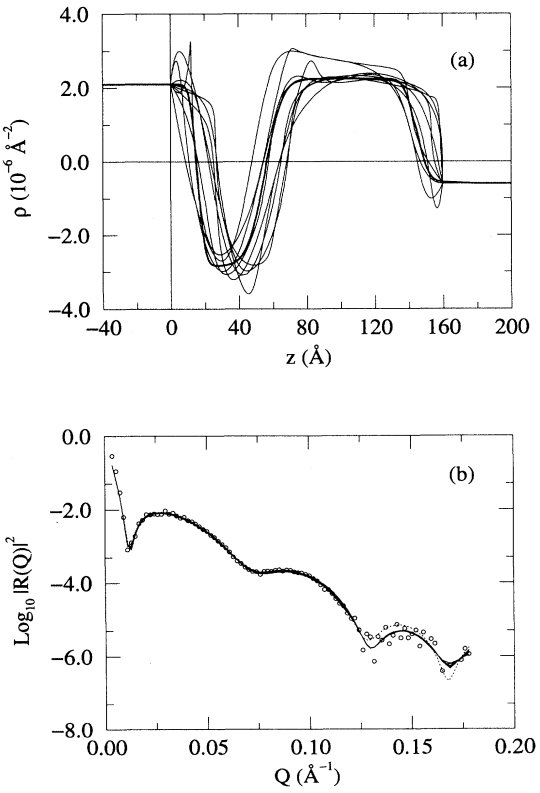

FIG. 6. More fits to data in Fig. 3. (a) Nine spline profiles resulting from various knot sequences, and the model fit from Ref. 40 (thick line). (b) Corresponding reflectivities (solid lines), and the model fit (dotted line) and data (circles) as in Fig. 3.

ored the training space without a need for policing. Of course, it was necessary to experiment with different sets of training configurations, including trials having different numbers of layers. We tried Oslo refinement in selected layers, but the uniform knot set produced equally good fits here. We have not attempted to improve this fit with a higher density of control points, since we believe that the neglect of instrumental resolution may limit the utility of further refinement of this spectrum. We note that while the spline multilayer profile is approximately periodic, there are small departures from perfect regularity which seem physically reasonable and are difficult to model using more highly constrained methods.

Finally, Fig. 8 shows a parametric B-spline fit to a simulated neutron reflectivity spectrum from a fictitious profile, which was generated from three piecewise continuous sinusoidal segments. The fit used 18 control vertices. The parameter knot sequence for the fit is shown in Table I. Note that the reflectivity falls ten orders of magnitude over the given Q range. The table indicates that the fit started with five control vertices and the uniform knot sequence $3,4,5,6,7$. The remaining knots were added over the course of the fit using Oslo refinement. The subdivisons of the $[3,4]$ segment reveal that the leading edge of the fictitious profile proved most difficult to fit in this strategy. In fact, this example proved to be difficult even knowing the goal profile. The successful strategy entailed first getting a close fit to only the low-Q portion of the spectrum with a relatively small number of segments, Fig. 9, and then using the "trickle" tactic described at the begining of this section to coerce the fit over the entire spectrum. Additional intervention was needed, however. The fit shown in Fig. 8 started with the one shown in Fig. 9, which used eight control vertices and a subset of the knot sequence in Table I, but it was necessary, ultimately, to forcibly translate the control graph in order to get the peak and the valley of the eventual fitted profile

TABLE I. Visible parameter knot sequence for the parametric B-spline fit shown in Fig. 8.

3.0	3.125	3.1875	3.25	3.375	3.5	3.75
4.0			4.25		4.5	4.75
5.0			5.25		5.5	5.75
6.0					6.5	
7.0						

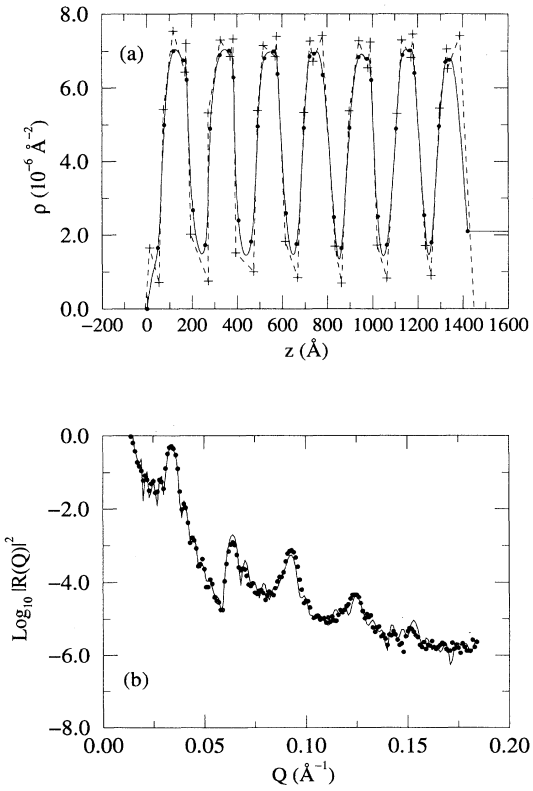

FIG. 7. Parametric B-spline fit of reflectivity from a seven-layer diblock copolymer multilayer on silicon, with air as the fronting medium. (a) Spline profile using 7×6 control vertices (41 segments). The phantom vertices are off scale. (b) Reflectivity data (points) and fit (line).

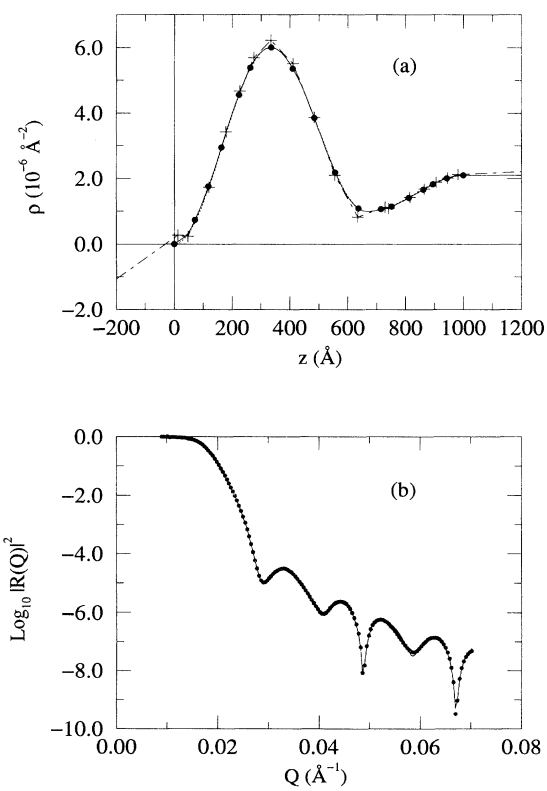

FIG. 8. Parametric B-spline fit of reflectivity from fictitious density profile. (a) A 17-segment spline fit (solid line) to the model profile made from 3 piecewise continuous cosine segments. The phantom vertices are far off scale. (b) Model reflectivity data (points) and fit (line).

to be able to align exactly with the goal. The locations of these features relative to the fixed end points (Sec. II C) are not secured by good fits to the truncated reflectivity spectrum because of trade-offs in the density profile shape that depend on the number of control vertices. One sees similar "slippage" in the titanium oxide fits of Fig. 6. Data at sufficiently high Q eventually may discriminate such features, but the relevant portion of the control space effectively may no longer be accessible to the searching algorithm. Figure 10 illustrates a complementary behavior. The fictitious goal profile is the solid line, while the dotted line is a manipulated version of the fitted profile in Fig. 8, obtained by moving control vertices by hand. The spline curve, while seemingly close to the model, actually weaves about it on a scale varying between approximately 50 Å and 200 Å. The corresponding reflectivity spectra break away from each other at large Q, where $|R(Q)|^2$ is very small for each, due to the sensitivity of large-Q scattering to $d\rho(z)/dz$ (Refs. 7, 9) and higher derivatives (Sec. II B). In this example, the derivatives of the piecewise cosine and parametric cubic B-spline curves of Fig. 10(a) are farther apart from each other than the profiles themselves are, especially along the leading edge and in the dip between 600 Å and 1000 Å. Of course, as Fig. 8 illustrates, B-spline profiles can be found that lie close to the model while also producing reasonably close fits of the spectrum in the indicated Q range. A fit procedure, naturally, will tend to find profiles that lie close to the goal in the reflectivity measure, even if—as the previous examples have illustrated—they do not always lie pointwise close in real space. However, as shown in Fig. 10, the "distance" between density profiles in the reflectivity measure also depends on their effective "smoothness," and density profiles that appear to be pointwise close do not necessarily produce the same reflectivities.

V. CONCLUSION

We have seen how parametric B-spline curves offer useful low-dimensional representations of scattering length density profiles in model-free analysis of specular reflectivity. The parametric representation affords an easy device for the depiction of relatively free-form profile shapes that may exhibit complicated features combining different degrees of smoothness. Furthermore, paramet-

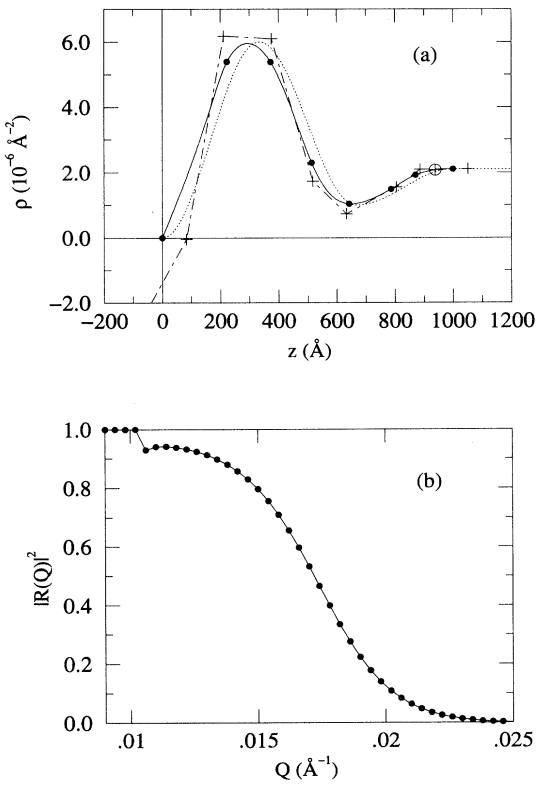

FIG. 9. Parametric B-spline fit of reflectivity from fictitious density in Fig. 7. (a) A seven-segment spline fit (solid line) to the model profile (dotted line) as determined by the portion of the refectivity shown in part (b) of this figure. Note "inboard" phantom at right. (b) Model reflectivity data (points) and fit (line) on a linear ordinate scale.

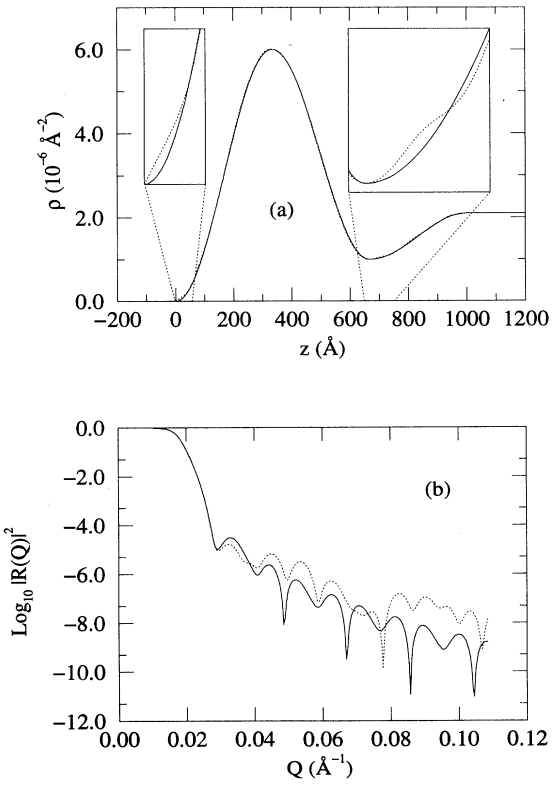

FIG. 10. (a) The piecewise cosine model profile of Fig. 7 (solid line) and a spline approximation (dotted line) made by manipulating the control vertices of the fitted spline profile of Fig. 7. The insets show magnified sections indicated by the dotted lines; on the left, 0–60 Å, on the right, 650–750 Å, which is typical of the range 600–1000 Å. (b) The corresponding reflectivities (solid and dotted lines, respectively).

ric curves are easy to render adaptively by binning the parameter u common to both the $z(u)$ and $\rho(u)$ components. Cubic B splines, in particular, are mathematically appropriate for profiles that are likely to be physically meaningful, and, in the curve fitting context, provide an effective balance between local and global control of profile shape. The Oslo algorithm permits dimensional refinement of B-spline descriptions while fits are in progress. Also, efficient B-spline computational procedures are available for practical implementation of these methods.

The Nelder-Mead simplex method is well suited to searching parametric B-spline spaces for fitting reflectivity spectra, but any optimization technique could be adapted to this purpose. The manifold property of the simplex algorithm affords passive enforcement of various constraints that are especially useful in these applications, including an effective means of honoring the restricted periodicity appropriate to the description of multilayers, as was discussed in Sec. IV, Fig. 7.

The use of parametric B-spline curves helps in the discovery of classes of profile shapes that are consistent with the data. As in the application to titaniun oxide discussed in Sec. IV, Figs. 4–6, and in the example of Fig. 9, given spectra may not be able to determine profiles in detail, but parametric B-spline curves can point to features that are common to classes of fitted shapes.

Based on these experiences, we remark briefly on the roles of model-free and model-based fitting procedures for analyzing reflectivities. While fitting to models is the more restrictive approach, such restriction may be what is desired when the goal is to test specific mathematical predictions of profile shape. Because function space is so large, it is difficult for model-free methods to find shapes having particular functional forms. For example, say the model fit shown in Figs. 4 and 6 is the model to be tested. None of the model-free curves found using parametric B splines fit this function exactly, and several not even closely, yet all reasonably fit the given spectrum. On the other hand, in model-based methods the models being used may be guided more by intuition than theory. This was the case in the titanium oxide model fit,[40] which is typical of many model-based approaches to reflectivity analysis, we believe. Then the concerns of testing the model are somewhat different, and it becomes helpful to have model-free methods available, as well, to put proposed models into context in regard to what profile shapes given spectra may support. Finally, of course, it may occur that no model easily can be found to explain a spectrum. Here model-free methods, such as the one we have described, become the tools of first choice, and the results may guide further refinement using models that are suggested by such analysis.

APPENDIX: PARAMETRIC B-SPLINE CURVES

1. Basics

B splines, $B_{i,k,\{u_i\}}(u)$, are non-negative, unimodular, piecewise functions of a variable u.[20] The pieces composing a B spline are kth-order polynomials, bounded by a discrete, ordered sequence of fixed u values, $\{u_i\}$, called knots. The index i locates the B spline on the knot sequence. For each i, $B_{i,k}(u)$ is composed of k polynomial segments on the contiguous intervals formed by the consecutive knots, $u_i, u_{i+1}, \ldots, u_{i+k}$, and are otherwise zero. For nondegenerate knot sequences, B splines of order k belong to continuity class C^{k-2}; this means they are continuous everywhere and have everywhere continuous derivatives, up to and including the $(k-2)$th, but generally not beyond. The class C^{-1} signifies piecewise discontinuous functions, which accommodates the $k=1$ case. The continuity requirement, along with normalization,

$$\sum_{j=0}^{k-1} B_{i-j,k}(u) = 1, \quad u_i \le u < u_{i+1}\,, \tag{A1}$$

completely determines $B_{i,k}(u)$ on a knot sequence.

Parametric B-spline curves are particular instances of parametric curves. The parametric representation of a curve means that its coordinates are functions of a real-valued parameter. A parametric curve $\mathbf{C}(u)$ in the x-y plane thus has the form

$$\mathbf{C}(u) = \big(x(u), y(u)\big)\,, \quad u_{\min} \le u \le u_{\max}\,. \tag{A2}$$

The curve is a trajectory in the plane, swept out as the variable u, called the evolution parameter, moves from $u_{\min}$ to $u_{\max}$. In the case of parametric B-spline curves, these coordinate functions are expressed as linear combinations of B splines,

$$\mathbf{C}(u) = \sum_{i=i_{\min}}^{i_{\max}} \mathbf{V}_i B_{i,k,\{u_i\}}(u)\,, \tag{A3}$$

where the fixed vectors

$$\mathbf{V}_i = (x_i, y_i), \quad i = i_{\min}, \ldots, i_{\max}\,, \tag{A4}$$

are called the control vertices of the curve.[20] These determine the curve for given order and knot sequence, since the B splines are unique, but generally are not on the curve itself.

For the curves of Eq. (A3) we use a standard correspondence between the control vertices and the knot sequence, which for given k is summarized by[20]

$$\begin{aligned} i_{\min} &= 0\,, \\ i_{\max} &= m \ge k-1\,, \\ \{u_i\}_0^{m+k} &= u_0 \le \cdots u_{k-1} \le \cdots \le u_{m+1} \le \cdots \le u_{m+k}\,, \\ u_{\min} &= u_{k-1} \quad (k\text{th knot from the left end})\,, \\ u_{\max} &= u_{m+1} \quad (k\text{th knot from the right end})\,. \end{aligned} \tag{A5}$$

A "standard" curve is defined by $m+1$ control vertices and is traced out as u moves from u_{k-1} to u_{m+1}. The $m-k+3$ parameter knots $u_{k-1}, \ldots, u_{m+1}$ are mapped onto the curve, subdividing it into $m-k+2$ segments. There are $m+k+1$ parameter knots, in all. The first $k-1$ parameter knots and the last $k-1$ parameter knots

are not mapped to the curve; their purpose is to provide a sufficient number of B-spline segments to ensure that each mapped parameter segment catches its full complement of k B-spline pieces. A standard curve is contained in the convex hull of the control graph. The control graph is the figure obtained by linearly interpolating the control vertices; its convex hull is bounded by the smallest convex polygon containing the control graph. In fact, the convex hull property applies to each segment individually. It follows that k collinear vertices produce a straight line segment, since their convex hull degenerates to a line.

$\mathbf{C}(u)$ is the union of its $m-k+2$ nonoverlapping segments, $\mathbf{C}_l(u)$, each of which is a weighted sum of k segment polynomials from k consecutive B splines:

$$\mathbf{C}(u) = \sum_{l=0}^{m-k+1} \mathbf{C}_l(u)\,. \tag{A6}$$

Segments of a standard curve are determined by overlapping sets of k control vertices, with $\mathbf{V}_l$ being the first in the sequence to have an effect on $\mathbf{C}_l(u)$, and $\mathbf{V}_{l+k-1}$ the last. Segments $\mathbf{C}_l(u)$ and $\mathbf{C}_{l'}(u)$ are independent of one another if $|l-l'|>k-1$, for then they have no control vertices in common. The parametric B-spline representation thus provides graded local control of the shape of curves, the range of control depending on the smoothness of the curve, in the sense of its continuity class. The more continuity that is required, the greater the order that must be used and the greater the segmental range of the B splines composing the curve. The smoother the curve is, in the precise sense of its continuity, the more extended the range of its control vertices; the less smooth it is, the more localized the control.

The continuity of parametric B-spline curves is not determined solely by the order k. Rather, C^{k-2} is the smoothest continuity class curves can be in. The actual continuity depends also on whether the knot or control sequences have degenerate entries. Degeneracies in either have similar geometrical consequences: The continuity class of a kth-order B spline having a μ-fold knot or vertex multplicity is $C^{k-\mu-1}$, where $\mu=1$ means nondegeneracy. In particular, in curves of order k, when $k-1$ consecutive vertices occupy the same point, they draw the curve into a corner cusp (discontinuous tangent) at that point.[20] It is noteworthy that such a simple device produces sharp features in an otherwise smooth (to the eye) curve, in contrast, say, to the use of Fourier composition. In the current application we use nondegenerate knot sequences but allow degenerate control vertices to occur in the fitting procedure.

2. Vector spaces and refinement

For a fixed knot sequence $\{u_i\}_0^{m+k}$, $B_{i,k,\{u_i\}}(u)|_0^m$ are a basis for the spline space $\mathbf{S}(\mathbf{P}^k,\{u_i\}_0^{m+k},C^{k-2})$ of dimension $m+1$. We abbreviate this as $\mathbf{S}^k$, or as $\mathbf{S}^k(\{u_i\}_0^{m+k})$ when needed. The fully articulated notation conveys that elements of $\mathbf{S}^k$ are formed piecewise from elements of $\mathbf{P}^k$, the space of kth-order polynomials, with C^{k-2} continuity at nondegenerate knots. Thus "splines," in general, may be defined as linear combinations of B splines.[18] The elements of $\mathbf{S}^k$ are the coordinate functions, and the basis $B_{i,k,\{u_i\}}(u)|_0^m$ can be lifted to the $2m$-dimensional direct product space $\mathbf{S}^k\otimes\mathbf{S}^k$, whose elements are the parametric spline curves $\mathbf{C}(u)$ in Eq. (A3).

Now consider that the knot sequence $\{u_i\}_0^{m+k}$ is contained in the sequence $\{w_j\}_0^{m+n+k}$, which is larger by n knots. This means that every u_i in the smaller set coincides with some w_j in the larger set. In practice, the n extra knots of $\{w_j\}_0^{m+n+k}$ are inserted into $\{u_i\}_0^{m+k}$, a process called knot refinement. Conventionally these new knots only go into the standard interval $[u_{k-1},u_{m+1})$ of the original sequence. Then $\mathbf{S}^k(\{u_i\}_0^{m+k}) \subset \mathbf{S}^k(\{w_j\}_0^{m+n+k})$; the original spline space is a vector subspace of the refined vector space. The refined space has a "new" B-spline basis $N_{j,k}(u) = B_{i,k,\{w_j\}}(u)$, which serves also a basis for any of its subspaces. Thus the original B splines can be expanded in the refined basis:

$$B_{i,k}(u) = \sum_{j=0}^{m+n} \alpha_{i,k}(j) N_{j,k}(u)\,. \tag{A7}$$

The transformation coefficients $\alpha_{i,k}(j)$ are called discrete B splines and can be calculated from the two knot sequences.[24]

A curve $\mathbf{C}(u)$ lifted from $\mathbf{S}^k(\{u_i\}_0^{m+k})$ by a control sequence $\{\mathbf{V}\}_0^m$ can also be lifted from an embedding space $\mathbf{S}^k(\{w_j\}_0^{m+n+k})$, using a refined control sequence $\{\mathbf{W}\}_0^{m+n}$:

$$\mathbf{C}(u) = \sum_{i=0}^{m} \mathbf{V}_i B_{i,k}(u) = \sum_{j=0}^{m+n} \mathbf{W}_j N_{j,k}(u)\,. \tag{A8}$$

Indeed, substitution of Eq. (A7) into Eq. (A8) leads to[20,24]

$$\mathbf{W}_j = \sum_{i=0}^{m} \alpha_{i,k}(j)\mathbf{V}_i\,, \quad j=0,\ldots,m+n\,. \tag{A9}$$

This method of spline curve refinement is known as the "Oslo algorithm"[24,25] and is incorporated into our fitting procedure.

3. Recurrence and implementation

On a given parameter knot sequence, B splines satisfy the recurrence[18,20]

$$B_{i,r}(u) = \frac{u-u_i}{u_{i+r-1}-u_i} B_{i,r-1}(u) + \frac{u_{i+r}-u}{u_{i+r}-u_{i+1}} B_{i+1,r-1}(u)\,, \tag{A10}$$

for $r=2,3,\ldots,k$. This is seeded by $B_{i,1}(u)$, which is unity for $u\in[u_i,u_{i+1})$, and zero otherwise. Degenerate knots are accommodated by ignoring terms with vanishing denominators. The curves $\mathbf{C}(u)$ can be computed

directly from a similar recurrence, found by substituting Eq. (A3) into Eq. (A10) to obtain[18,20]

$$\mathbf{V}_i^{[r-1]}(u)=\frac{u-u_i}{u_{i+r-1}-u_i}\mathbf{V}_i^{[r]}(u)+\frac{u_{i+r-1}-u}{u_{i+r-1}-u_i}\mathbf{V}_{i-1}^{[r]}(u)\,, \tag{A11}$$

for $r = k,\ldots,2$. When Eq. (A11) is initialized with $\mathbf{V}_i^{[k]}(u) = \mathbf{V}_i$ for $i = 1,\ldots,m$, it terminates with

$$\mathbf{C}(u) = \mathbf{V}_\delta^{[1]}(u)\,, \tag{A12}$$

where δ is the unique index such that $u_\delta \le u < u_{\delta+1}$. The pseudocode for this procedure is given in Ref. 20, Chap. 20. Actually, Eq. (A12) is computationally expensive where fine-grained rendering is needed. In our calculations, we use, instead, a more efficient variation of this scheme, prescribed for situations in which several values of u are likely to trigger the same value of δ.[44] A similar recursive procedure can be developed for the Oslo algorithm, Eq. (A9). For that we use a version of the pseudocode in Ref. 45.

4. Parameter knot selection

While the number of parameter knots in the sequence $\{u_i\}_0^{m+k}$ is fixed by the number m of control vertices used to make $\mathbf{C}(u)$, the actual location of these knots remains to be chosen. The choice can be guided by the convenient fact that B splines are invariant under affine transformations of the u axis. Let $u = a\bar{u} + b$, for any constants a and b. Then

$$B_{i,k,\{u\}}(u) = B_{i,k,\{\bar{u}\}}(\bar{u})\,, \tag{A13}$$

for $u_0 \le u < u_{m+k}$ and for $\bar{u}_0 \le \bar{u} < \bar{u}_{m+k}$. This is seen by substituting

$$\frac{u-u_i}{u_{i'}-u_{i''}} = \frac{\bar{u}-\bar{u}_i}{\bar{u}_{i'}-\bar{u}_{i''}}$$

in Eq. (A10), for any i, i' and i''. Thus, from Eq. (A3), it follows that, for a given control graph,

$$\mathbf{C}(u) = \mathbf{C}(\bar{u})\,. \tag{A14}$$

Uniform B splines are composed on uniform knot sequences, $\{u_i\}_0^{m+k} = u_0 + i\Delta$, for any constant Δ. They have the unique property of being translates of each other:

$$B_{i,k}(u) = B_{0,k}(u - i\Delta)\,.$$

Because of affine invariance, uniform knot sequences can be restricted to the integer sequences[20]

$$0, 1, \ldots, k-1, \ldots, m+1, \ldots, m+k. \tag{A15}$$

These are convenient starting points for adding knot degeneracies and for Oslo refinement, both of which destroy uniformity. In our reflectivity analysis, we will always begin with uniform sequences and, if desired, subdivide intervals by halving for Oslo refinement.

[1] Proceedings of the Workshop on "Methods of Analysis and Interpretation of Neutron Reflectivity Data," edited by G. P. Felcher and T. P. Russell, [Physica B **173**, 1 (1991)].

[2] T. M. Roberts, Physica B **173**, 157 (1991).

[3] S. K. Sinha, M. K. Sanyal, K. G. Huang, A. Gibaud, M. Rafailovich, J. Sokolov, X. Zhao, and W. Zhao, in *Surface X-Ray and Neutron Scattering*, edited by H. Zabel and I. K. Robinson (Springer-Verlag, Berlin, 1992), p. 85.

[4] D. S. Sivia, W. A. Hamilton, G. S. Smith, T. P. Rieker, and R. Pynn, J. Appl. Phys. **70**, 732 (1991).

[5] D. S. Sivia, W. A. Hamilton, and G. S. Smith, Physica B **173**, 121 (1991).

[6] J. Penfold and R. K. Thomas, J. Phys. Condens. Matter **2**, 1369 (1990).

[7] J. S. Pedersen, J. Appl. Crystallogr. **25** 129 (1992).

[8] N. Singh, M. Tirrel, and F. S. Bates, J. Appl. Crystallogr. **26**, 650 (1993).

[9] X.-L. Zhou and S.-H. Chen, Phys. Rev. E **47**, 3174 (1993).

[10] V.-O. de Haan and G. G. Drijkoningen, Physica B **198**, 24 (1994).

[11] K. Kunz, J. Reiter, A. Götzelmann, and M. Stamm, Macromolecules **26**, 4316 (1993).

[12] J. S. Pedersen and I. W. Hamley, Physica B **198**, 16 (1994).

[13] I. W. Hamley and J. S. Pedersen, J. Appl. Crystallogr. **27**, 29 (1994).

[14] J. S. Pedersen and I. W. Hamley, J. Appl. Crystallogr. **27**, 36 (1994).

[15] The rendering step is unnecessary in the Born approximation if the Fourier transforms of the bases functions are known exactly, as in Refs. 7, 8, 12–14. Use of the Born approximation, however, entails an assumption that is not always correct and needs to be tested for each spectrum where low-Q data are available.

[16] O. Glatter, J. Appl. Crystallogr. **10**, 415 (1977).

[17] A version of this work first was presented at *New Horizons: A Workshop on the State of the Art in Neutron Reflectivity*, National Institute of Standards and Technology, Gaithersburg, MD, December 9-10, 1993 (unpublished).

[18] C. de Boor, *A Practical Guide to Splines* (Springer-Verlag, New York, 1978).

[19] L. L. Shumaker, *Spline Functions: Basic Theory* (Wiley, New York, 1981).

[20] R. H. Bartels, J. C. Beatty, and B. A. Barsky, *An Introduction to Splines for Use in Computer Graphics and Geometric Modeling* (Morgan Kaufmann, Los Altos, CA, 1987).

[21] In general the $\rho(z)$ are complex-valued functions, in order to account both for scattering (real part) and absorption (imaginary part). Thus in the parametric B-spline representation, the $\{\rho_i\}$, and hence, $\rho(u)$, are complex also. We leave this implicit, but it means that the function space of concern actually is larger than the notation suggests. For neutrons, absorption is negligible in most cases, and the assumption of real-valued $\rho(z)$ is a good approximation. For x rays, the imaginary part of $\rho(z)$ often is important, and the

fitting procedure then must allow for this additional freedom. In fact the methods we describe are not restricted to real-valued density profiles, but effective fitting strategies should be informed of the need to find complex-valued profiles when they are required.

[22] Reentrant "$\rho(z)$" do not present computational problems, *per se*. The various methods of calculating reflectivities from density profiles need only to associate a ρ value with every z value, without necessary regard for the possibility of recurring z. The parametric representation accomplishes this, since every u value is mapped to only one (z, ρ) coordinate, regardless of where the point falls. Inappropriate z recurrences are algorithmically innocuous. The resulting reflectivities are "correct" for the given density profile, even if they are not physically meaningful.

[23] Reference 20, Sec. 20.3.

[24] E. Cohen, T. Lyche, and R. Riesenfeld, Comput. Graphics Image Process. **14**, 87 (1980).

[25] H. Prautzsch, Comput. Aided Geom. Design **2**, 329 (1985).

[26] References 12–14 employ an informal notion of smoothness as a least-squares fitting constraint.

[27] M. J. Lighthill, *An Introduction to Fourier Analysis and Generalized Functions* (Cambridge University Press, New York, 1958).

[28] S. K. Sinha, E. B. Sirota, S. Garoff, and H. B. Stanley, Phys. Rev. B **38**, 2297 (1988).

[29] C. F. Majkrzak, N. F. Berk, S. K. Satija, and T. P. Russel, Proc. SPIE **1738**, 282 (1992).

[30] B. A. Barsky, Comput. Ind. **3**, 17 (1982).

[31] J. A. Nelder and R. Mead, Comput. J. **7**, 308 (1965).

[32] S. L. S. Jacoby, J. S. Kowalik, and J. T. Pizzo, *Iterative Methods for Nonlinear Optimization Problems* (Prentice-Hall, Englewood Cliffs, NJ, 1972). See pp. 79–83 for a lucid definition of the Nelder-Mead algorithm, including a detailed flow diagram.

[33] See also W. M. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes* (Cambridge University Press, Cambridge, England, 1990).

[34] If $a_1 = b_1$ and $a_2 = b_2$, then $\alpha a_1 + \beta a_2 = \alpha b_1 + \beta b_2$ for all α, β. This, of course, is the generalization of the rule, "equals added to equals are equal."

[35] If $a_1 \geq b_1$ and $a_2 \geq b_2$, then $\alpha a_1 + \beta a_2 \geq \alpha b_1 + \beta b_2$ when $\alpha, \beta \geq 0$; but otherwise, not always.

[36] Fitting E to D on a logarithmic scale accomplishes a similar goal, of course, but is more expensive computationally and produces flatter objectives, which slow the movement of the simplex.

[37] P. Croce and B. Pardo, Nouv. Rev. Opt. Appl. **1**, 229 (1970).

[38] S. Yamada, T. Ebisawa, N. Achiwa, T. Akiyoshi, and S. Okamoto, Annu. Rep. Res. Reactor Inst. Kyoto Univ. **11**, 8 (1978).

[39] See Ref. 9 for a different solution of this problem.

[40] D. G. Wiesler and C. F. Majkrzak, Physica B **198**, 181 (1994).

[41] The scattering length density for the backing medium has a small effective imaginary part (due to isotropic incoherent scattering), $\mathrm{Im}\rho_b = 0.015 \times 10^{-6}$ Å^{-2}, which has negligible effect on the fits over the given Q range.

[42] J. F. Ankner and C. F. Majkrzak, Proc. SPIE **1738**, 260 (1992).

[43] S. K. Satija and T. P. Russell (private communication).

[44] Reference 20, Secs. 20.2, 20.4, 20.5.

[45] Reference 20, Sec. 20.6.